



**STANDARDS REFORM, FEDERAL LAW,  
AND THE AMERICAN DIPLOMA PROJECT:**

**A FRAMEWORK FOR MAKING LEGALLY SOUND DECISIONS**

**Arthur L. Coleman  
Scott R. Palmer**



**NIXON PEABODY LLP**  
ATTORNEYS AT LAW

## ABOUT THE AUTHORS

Arthur Coleman and Scott Palmer are Counsel at Nixon Peabody LLP in Washington, D.C., where they focus on preventive law in education – assisting states and school districts involved in standards-based reforms develop policies and programs that best serve their education goals and are consistent with federal and state legal standards, thereby reducing the risk of litigation or enforcement. Mr. Coleman and Mr. Palmer each served, back-to-back, as Deputy Assistant Secretary for Civil Rights in the U.S. Department of Education during the Clinton Administration. During their tenure, they led the Department's effort to produce *The Use of Tests as Part of High-Stakes Decision-Making for Students: A Resource Guide for Educators and Policy-Makers* (2000).

Mr. Coleman is a member of the National Academy of Sciences Board on Testing and Assessment's Committee on Educational Excellence and Testing Equity, and serves on the Editorial Board of the *Education Assessment Insider: Compliance, Testing and Accountability*. He is an honors graduate of Duke University School of Law.

Mr. Palmer is Senior Legal Fellow at the Harvard University Civil Rights Project and Adjunct Professor at the University of Maryland, where he teaches courses on education law and policy. He is a graduate of University of California, Berkeley, School of Law, Boalt Hall, and received a Masters in Public Policy from Harvard University's John F. Kennedy School of Government.

## TABLE OF CONTENTS

INTRODUCTION .....	1
I. FEDERAL LAW RELATED TO THE ADP.....	4
A. OVERVIEW OF RELEVANT FEDERAL LAWS .....	4
1. LEGAL PRINCIPLES IN EDUCATION.....	4
2. LEGAL PRINCIPLES IN EMPLOYMENT .....	8
B. COMPARING K-12 EDUCATION, HIGHER EDUCATION, AND EMPLOYMENT	11
II. THE AMERICAN DIPLOMA PROJECT AND FEDERAL LAW.....	13
A. THREE WAYS THE ADP IMPLICATES FEDERAL LAW .....	13
III. A FRAMEWORK FOR IDENTIFICATION AND ANALYSIS OF CENTRAL ISSUES	
.....	16
A. PURPOSE.....	18
B. VALIDITY.....	19
C. ADMINISTRATION AND USE.....	25
D. OPPORTUNITY TO LEARN.....	27
E. LESSONS LEARNED.....	29
IV. THE PROCESS FOR MOVING FORWARD .....	32

## INTRODUCTION

The standards reform movement seeks primarily to strengthen accountability in K-12 education to ensure that all students achieve high standards. With the recent enactment of the federal No Child Left Behind Act and ongoing efforts in every state, standards reform is rapidly expanding and holds great promise for students across the country. As with most significant policy developments, however, standards reform has also raised a new generation of legal issues and challenges under federal law.

The American Diploma Project (the “ADP”) seeks to strengthen standards-reform efforts in five states, and ultimately nationwide, by (1) fully aligning state high school standards in reading, writing, and mathematics with college admissions and employment standards, (2) encouraging colleges and employers to use state assessment data in their admissions and hiring decisions, and (3) establishing new benchmarks to assist with similar efforts in other states. In so doing, the ADP, too, implicates federal legal issues.

The fact that new legal issues surface in the context of the ADP, however, does not suggest that the legal risks outweigh the educational or other benefits of the project. To the contrary, properly understood, federal civil rights and related laws reinforce sound educational practices. Far from threatening standards reform or the ADP, these legal standards – if used to guide and shape the implementation of the project on the front end – can help provide an important foundation for ensuring that the ADP is implemented in ways that best serve its goals and are legally sustainable.

This paper provides an overview of the major federal legal issues and standards implicated by the ADP, and provides a framework to help participating states – including their K-12, higher education, and business communities – move forward in a manner that minimizes legal risk. The first section briefly discusses the federal laws relevant to the ADP.<sup>1</sup> The second section explains the several

---

<sup>1</sup> This paper addresses the federal laws and principles most implicated by the design and projected state implementation of the ADP: due process and nondiscrimination. See generally U.S. Dept. of Educ. Office for Civil Rights, *The Use of Tests as Part of High-Stakes Decision-Making for Students: A Resource Guide for Educators and Policy-Makers* (2000) [“OCR Resource Guide”]; National Research Council, *High Stakes: Testing for Tracking, Promotion, and Graduation* (1999) [“NRC High Stakes”]. There are additional issues of federal law related to more narrow aspects of ADP implementation that may also arise, such as student privacy issues under the Family Educational Records and Privacy Act [“FERPA”], 20 U.S.C. §§ 2701 *et seq.*, which may surface upon the transfer or disclosure of student education records, and academic freedom issues under the First Amendment to the U.S. Constitution, which may surface in higher education in debates regarding “who has authority” to set standards for admissions or placement. These issues are not addressed in this paper.

distinct ways that the ADP likely implicates federal law. The third section provides a framework for analyzing ADP efforts in light of federal law and identifies several ADP-specific issues and lessons. The final section lists next steps that should be taken by each participating state in implementing the ADP.<sup>2</sup>

The federal legal issues raised by the ADP are grounded in decades-old principles of federal law, and there is a lot that we know regarding those principles and how they likely apply to the ADP. However, in many areas, relevant federal caselaw is sparse or even nonexistent, and in some areas, the ADP differs from historical antecedents in significant ways. This fact – coupled with the fact that the ADP will be implemented in different ways in different state contexts – means that this paper cannot and does not provide absolutes (as with any prospective legal analysis); rather, this paper identifies and analyzes key legal principles in the context of the broad ADP design and makes the best possible judgments to help inform development and implementation of the project.

Importantly, nothing in this paper or our analysis suggests a “red light” with regard to the ADP, but this paper is not a “green light” either. Rather, this paper reflects a strong “yellow light” that can be moved toward “green” with appropriate implementation (or could become “red” without it). Moreover, there is no simple formula or checklist that will guarantee compliance. Rather (and particularly at this level of generality and at this stage of the ADP process), it is better to view potential legal risk on a spectrum (much like conclusions of test validity, in fact) – where there are few categorical rules but where there are numerous factors that can help establish that a program or practice is and would likely be found to be educationally and legally sound.

Although this paper should not be read as a definitive predictor of the law as it may apply to some of the novel issues raised by the ADP, it has been written with an eye toward potential claims and ways to move forward most appropriately in light of those potential claims. In that sense, the objective of this overview is two-fold:

---

<sup>2</sup> Although cast in legal terms with specific reference to a legal framework of questions, this document has been prepared with the intention that it be accessible to ADP participants from multiple backgrounds. As a result, it does not seek to describe all the nuances of relevant laws or cases in great detail (not that such an effort would be possible or practical at this stage of the ADP). Rather, this document reflects an effort to identify, distill, and discuss the most germane principles relevant to the unique facets of the ADP in a way that will have resonance for all ADP participants. To the extent that this document reads in some places more like a synthesis of psychometric, rather than legal, principles, it merely reflects the centrality of test measurement principles to any federal legal analysis. See *generally* *OCR Resource Guide*; *NRC High Stakes* (“legal and psychometric standards reflect many common concerns”). That being said, this discussion attempts to cast the relevant issues in legal terms, as the courts have viewed them. Thus, there are areas in which this paper will not reflect perfect symmetry with psychometric terminology or concepts.

(1) to provide meaningful information to ADP partners so that they can work to avoid facing legal claims at the outset; and (2) to help them develop the foundations to successfully defend against those claims should they arise. By so doing, this analysis can also help ADP partners improve educational opportunities and outcomes for all students. It goes without saying that part of the analysis also includes an examination of areas in which the unique features of the ADP will likely support and enhance its lawfulness.

Finally, many of the federal legal issues raised by the ADP are similar across states, and this document can therefore serve as a foundation for all state participants. However, state context and decisions by each state's K-12, higher education, and business communities matter greatly to any specific legal analysis. In almost every federal case regarding standards reform and other issues relevant to the ADP, the context and facts surrounding the challenged practice have driven the ultimate conclusion reached by the federal court or enforcement agency. Said differently, specific facts drive legal results.<sup>3</sup> Just as federal laws can reinforce sound educational (or business) practices, so do state educational decisions drive the legal analysis and form the basis of evidence necessary to defend a given program under federal law. This paper does not address unique issues of state policy or law that may arise in the ADP context, and it should not be read as state-specific legal advice. However, it does provide guidance that is designed to help foster discussion and analysis regarding state-specific issues. The logical next step is for each ADP state to apply this guidance – including the framework, lessons learned, and process described herein – to its specific state context, and to develop a state-specific plan to ensure that the key legal issues identified here are addressed throughout the ADP's development and implementation process.

---

<sup>3</sup> Illustrating this point, one federal court within the span of three years addressed opportunity to learn and discrimination issues that surfaced in the context of a state minimum competency exam required for high school graduation, concluding that the necessary foundations that would permit the state to deny a diploma to students who had failed the statewide competency exam were not in place in 1979, but that, with policy changes accompanied by additional educational supports and interventions, the state could deny a diploma to students who repeatedly failed the exam beginning with the class of 1983. Compare *Debra P. v. Turlington*, 644 F. 2d 397 (5<sup>th</sup> Cir. 1981) [*Debra P.*], with *Debra P. v. Turlington*, 730 F. 2d 1405 (11<sup>th</sup> Cir. 1984).

## I. FEDERAL LAW RELATED TO THE ADP

Whenever an actor (e.g., a state, college, or employer) makes decisions that limit or deny educational or other benefits to some individuals (e.g., graduation, admissions, or employment decisions), those decisions must adhere to certain federal laws. Precisely what federal laws apply and how they apply will depend on several factors, including: (1) who the actor is (e.g., public or private); (2) the actor's purpose (i.e., the nature of the decision); (3) what factor(s) is used in making the given decision (e.g., a test score or high school diploma); (4) whether that factor is valid for the given purpose; (5) how that factor is used; and more. This section briefly discusses key federal laws related to standards reform and the use of assessments as a foundation for analyzing the likely implications of those laws for the ADP. Although it is beyond the scope of this document to synthesize every case of potential relevance to the ADP, it is important to understand the application of these broad federal principles to have a better contextual understanding of the legal baselines and to have a better foundation for addressing the kinds of legal issues that may arise as a result of the ADP.

### Overview of Relevant Federal Laws

Amidst a panoply of federal statutes and regulations that touch on decisions that affect education and employment opportunities, there are two overarching federal doctrines that are most applicable to the ADP: (1) due process and (2) nondiscrimination. These legal issues most often arise in cases where decisions are made affecting the provision of educational or other benefits to individuals (as opposed to decisions regarding school- or district-level performance).<sup>4</sup> Simply put, those decisions must in general be nondiscriminatory and consistent with due process, as described below. Federal due process protections apply only to public entities, such as states, school districts, and public universities, while federal nondiscrimination laws, through various mechanisms, may apply to both public and private actors, including states, universities, and employers. Although these doctrines operate from different vantage points, they both ultimately hinge on the notion of “fairness” and the question of whether valid educational or business decisions affecting students or employees are being made.

***Legal Principles in Education.*** Each of the two core legal doctrines, due process and nondiscrimination, may arise in the education context, and they have most frequently surfaced with regard to K-12 education. First, under the 14<sup>th</sup> Amendment to the U.S. Constitution, states and other public actors must afford *due process* whenever they act to limit or deny individuals a liberty or property

---

<sup>4</sup> The focus of the ADP is on student accountability measures, rather than system accountability measures, which this paper does not address.

interest, which may include certain benefits that a person has a reasonable expectation of receiving. In applying due process in the educational setting, federal courts have generally focused on three questions:

- Was the student denied a property or liberty interest sufficient to implicate federal due process protections (such as the denial of a high school diploma)?
- If so, is the challenged action reasonably related to a legitimate governmental interest (such as raising learning standards)?
- If so, have students been provided with sufficient notice and a fair opportunity to learn the material being tested?<sup>5</sup>

For example, several due process claims have arisen in cases where a student was denied a high school diploma for failing to pass a state high school exit exam. In these cases, courts have generally held that a diploma constitutes a sufficient property interest to implicate due process.<sup>6</sup> In examining whether the exit exam requirement is reasonably related to a legitimate interest, courts have generally given substantial deference to the state regarding the legitimacy of the educational ends to be achieved and looked primarily for evidence that the given exam is in fact a valid means for achieving those ends. Finally, courts have examined whether students had sufficient notice of the graduation requirement and whether the curriculum and instruction were generally aligned with the content of the exam such that students had a meaningful opportunity to learn the material being tested.

Second, several federal laws, such as Title VI of the Civil Rights Act of 1964, prohibit *discrimination* by recipients of federal funds (which include all states and school districts, all public universities, nearly all private universities, and more) on the basis of race, color, and national origin, as well as gender or disability.<sup>7</sup> These

---

<sup>5</sup> The area of public employment raises comparable issues but is not addressed in this paper because those issues are not central to the ADP design at this time. See, e.g., *Allen v. Alabama State Board of Education*, 976 F. Supp. 1410 (M.D. Ala. 1997) (state obliged to comply with consent decree requiring that state's teacher test be matched to curriculum of state's teacher education programs unless state demonstrates that it does not control teacher education). See generally Diana Pullin, *Key Questions in Implementing Teacher Testing and Licensing*, 30 J.L. & Educ. 383 (2001).

<sup>6</sup> The question of whether a high school diploma implicates a property right sufficient to trigger due process may not be subject to a categorical conclusion. Compare *Debra P. with Bester v. Tuscaloosa City Board of Education*, 722 F.2d 1514 (11<sup>th</sup> Cir. 1984) (distinguishing *Debra P.*'s ruling that a diploma denial implicates due process rights).

<sup>7</sup> The federal laws most directly implicated in this analysis include: Title VI of the Civil Rights Act of 1964, which prohibits discrimination on the basis of race, color, or national origin by recipients of federal funds; Title IX of the Education Amendments of 1972, which prohibits



statutes, as well as the 14<sup>th</sup> Amendment, generally prohibit intentional discrimination or the “different treatment” of persons based on race, national origin, or other suspect classifications. In addition, and likely of greater applicability in the ADP context, these statutes and/or their implementing regulations also prohibit policies or programs that are neutral on their face but have the effect of discriminating in that they have an unjustified “disparate impact” by race, national origin, or other protected classification.<sup>8</sup> Importantly, disparity alone does not equal discrimination. Rather, the existence of a significant disparity by race, national origin, or other protected classification in the distribution of benefits triggers the need for further inquiry to ensure that the policy causing the disparity is in fact nondiscriminatory. The disparate impact standard, in the education context and more broadly, thus consists of a three-part test, which is designed to distinguish discrimination from disparity alone.

- Does the policy result in a significant disparity in the distribution of benefits by race or national origin (for example)?
- If so, is the policy educationally justified?
- If so, is there an alternative policy that would equally serve the institution’s goals with lesser disparity?

For example, some Title VI disparate impact cases have arisen in the context of state high school exit exam requirements. In such cases, the plaintiff (e.g., the student) has the burden of showing that the exit exam results in a significant disparity (in both statistical and practical terms) in the award of benefits – in this case a high school diploma – by race or national origin. If the plaintiff makes this showing, then the defendant (e.g., the state) must offer an educational justification for the policy. As in the due process analysis, courts will generally defer to the state regarding the educational ends to be achieved, as long as they are legitimate, and look for evidence that the given exam is in fact a valid way to achieve those ends. If the defendant demonstrates the “educational necessity” of

---

discrimination on the basis of gender by recipients of federal education funds; Section 504 of the Rehabilitation Act of 1973, which prohibits discrimination based on disability by recipients of federal funds; Title II of the Americans with Disabilities Act of 1990, which prohibits discrimination based on disability by public entities; and more.

<sup>8</sup> In the recent case of *Alexander v. Sandoval*, 532 U.S. 275 (2001), the U.S. Supreme Court ruled that the Title VI disparate impact regulations do not afford private litigants a remedy in federal court. Nonetheless, this decision does not effect the jurisdiction or authority of the U.S. Department of Justice or the U.S. Department of Education’s Office for Civil Rights [“OCR”] to enforce these regulations. Moreover, some courts have held since *Sandoval* that a private right of action alleging disparate impact discrimination against public actors may be sustained pursuant to 42 U.S.C. §1983, which prohibits any person acting under color of state law from violating federal laws. See, e.g., *White v. Engler*, No. 00-CV-72882-DT, slip op. (E.D. Mich. Nov. 19, 2001). But see *S. Camden Citizens in Action v. N. J. Dep’t. of Env’tl. Prot.*, 274 F.3d 771 (3<sup>rd</sup> Cir. 2001).

the exit exam, then, to prevail in the case, the plaintiff must demonstrate that there are other policies available that would equally serve the institution's goals with lesser disparity.

To illustrate these due process and nondiscrimination standards, consider the recent case of *GI Forum v. Texas Education Agency*,<sup>9</sup> in which the State of Texas successfully defended its high school exit exam against both due process and discrimination claims. Although representing the view of only one federal judge, the case presents a reasoned analysis that tracks the principles discussed above.<sup>10</sup>

The court in *GI Forum* ruled that the Texas high school exit exam [“the TAAS”] did not discriminate against minority students in violation of the Title VI disparate impact standard, and did not deny plaintiffs their due process rights guaranteed under the Fourteenth Amendment. On the Title VI claim, the court found that the disparities between the passing rates of minority and majority students were statistically significant so as to trigger further inquiry. Despite the “large,” “disconcerting,” and “sobering differences” in pass rates, however, the court found that there was a “manifest relationship” between the State’s educational goals and use of the TAAS. Among other things, the test provided an “objective assessment” of mastery of a discrete set of skills and knowledge linked to State educational standards, served to motivate students, and provided support for “systemic accountability.” Finally, the court found that the TAAS served these goals better than any alternative presented to the court. As a consequence, the court concluded that plaintiffs failed to sustain their burden of proving a Title VI violation, despite the disparate impact against racial minorities present in the case. On the due process claim, the court concluded that the TAAS met “currently accepted standards for curricular validity” (measuring what it was supposed to measure), and that “all students” in Texas had a “reasonable opportunity to learn the matter covered by the test” – particularly given the State’s remediation efforts and the fact that students had multiple opportunities to pass the test.<sup>11</sup> Central to the court’s reasoning on the due process and discrimination claims was its belief that the TAAS was ultimately educationally beneficial to all students, and that the State was using assessment data to address, rather than exacerbate, disparities in educational opportunity and achievement.

Finally, a number of federal statutes require that states include students with disabilities and English language learners in state testing programs and that states

---

<sup>9</sup> *GI Forum v. Texas Education Agency*, 87 F. Supp. 2d 667 (W.D. Tex. 2000).

<sup>10</sup> The court applied the principles from *Debra P.* in reaching its conclusion, and the standards articulated are consistent with the OCR *Resource Guide* as well.

<sup>11</sup> *GI Forum*, 87 F. Supp. 2d at 672, 682, 683.

provide them with appropriate accommodations. For example, the Individuals with Disabilities Education Act, Section 504 of the Rehabilitation Act, and the Americans with Disabilities Act, prohibit discrimination based on disability and, among other things, require the inclusion of students with disabilities in state testing programs, the provision of appropriate accommodations, and more. Here too, the central issue concerns validity – whether the given assessment, with appropriate accommodations, in fact measures what it is supposed to measure (i.e., what a student knows or is able to do) in the given context, rather than measuring irrelevant constructs related to the student’s disability or limited English proficiency (unless those constructs are what the test is intended to measure). Much of the initial wave of litigation and enforcement activity related to standards reform has centered on claims by disabled students raising issues related to accommodations, sometimes regardless of the attachment of high-stakes consequences.<sup>12</sup>

**Legal Principles in Employment.** Though due process issues can arise in the context of public employment, most employment cases involve nondiscrimination principles, which are similar to those discussed above. In the employment context, several federal laws, most notably Title VII of the Civil Rights Act of 1964, prohibit discrimination in employment based on race, national origin, and other protected classifications. Title VII expressly includes in the statute itself a prohibition against employment practices that have an unjustified disparate impact by race or national origin (for example). The central questions in the employment setting are (1) whether the given employment policy causes significant disparities in benefits by race or national origin, (2) if so, whether the given practice is “job related for the position in question and consistent with business necessity,”<sup>13</sup> and (3) if so, whether less discriminatory alternatives exist.<sup>14</sup>

---

<sup>12</sup> See, e.g., *Rene v. Reed*, 751 N.E.2d 736 (Ind. Ct. App. 2001); *Parents for Educational Justice v. Picard*, 2000 LEXIS 6382 (E.D. La. 2000) *aff’d mem. sub nom* 2001 U.S. App. LEXIS 23099 (5<sup>th</sup> Cir. 2001); *Florida State Dept. of Education*, 28 IDELR 1002 (OCR 1998); *Virginia Dept. of Education*, 27 IDELR 1148 (OCR 1997); *Nevada State Dept. of Education*, 25 IDELR 752 (OCR 1996); *Hawaii Dept. of Education*, 17 EHLR 360 (OCR 1990); *South Carolina Dept. of Education*, 352 EDLR 475 (OCR 1987);

In a recent preliminary ruling that is subject to appeal, a federal district court in California refused to halt the administration of the State’s high school exit exam, which had been challenged by students with disabilities on multiple grounds, but ruled that the right to “meaningful inclusion” under the IDEA provided the necessary justification (irrespective of high-stakes consequences associated with the test) for the court to exercise jurisdiction and to direct the modification of certain state policies related to accommodations, modifications, and alternate assessments for students with disabilities. *Chapman v. California Dept. of Education*, Case No. G-01101780/JCS (N.D. Cal. 2002)

<sup>13</sup> 42 U.S.C. §2000e-2(k)(1)(A)(i).

<sup>14</sup> This employment standard served as the foundation for the development of the disparate impact standard in education matters. Despite conceptual similarities between the standards, there

Similar to the education context, court decisions in the employment context require some appropriate level of evidence demonstrating the validity of the employment selection measure for the given employment decision. Selection criteria, in the words of the U.S. Supreme Court, must be “demonstrably a reasonable measure of job performance.”<sup>15</sup> In practical terms, this means that the selection criteria must bear a substantial relation to the selection of persons with the skills and abilities necessary to do the job in question.<sup>16</sup> Ranging from the most technical and complex validation study to the more general research and opinion testimony supporting employers’ efforts, courts have considered a variety of kinds of validation evidence when evaluating Title VII claims.

Illustrating these principles in the testing context, and the complexity of the legal task, is the recent case of *Association of Mexican American Educators v. State of California*,<sup>17</sup> in which a divided full panel of the Ninth Circuit Court of Appeals ruled that the State of California’s teacher certification test was not discriminatory under Title VI and Title VII because its was sufficiently validated. Challenged in part because the test was not validated with specific reference to the *particular* jobs for which it was used – ranging from “twelfth grade mathematics teachers to seventh

---

are notable distinctions, such as the fact that Title VI applies only to recipients of federal funds and Title VII applies to any public or private employer that employs fifteen or more employees.

<sup>15</sup> *Griggs v. Duke Power*, 401 U.S. 424, 436 (1971).

<sup>16</sup> Among many courts, great deference is afforded to the Equal Employment Opportunity Commission’s 1978 *Uniform Guidelines on Employee Selection Procedures* [“*Uniform Guidelines*”], which establish three acceptable types of validation studies: criterion-related, content-related, and construct validation studies. 29 C.F.R. Part 1607. However, there are a number of ways in which employers may defend their use of selection criteria as valid and nondiscriminatory – some of which do not have to track the *Uniform Guidelines*.

The particular categories of test use validity, such as those reflected in the *Uniform Guidelines*, are not described further in this document. Despite the recent publication of the *Standards for Educational and Psychological Testing* (1999) by the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education [hereinafter *Joint Standards*], in which construct validity is discussed as the “umbrella” category that encompasses, for example, predictive, content, and criterion validity, OCR *Resource Guide* at 24, n.77, there has been (and is) no general agreement about specific terminology. (In fact, in *GI Forum*, the court found sufficient “curricular validity,” but the expert who later reported on the case characterized the validity in issue as one of “content validity.” See *GI Forum*, 677 F. Supp. 2d at 682; Phillips, *GI Forum v. Texas Education Agency: Psychometric Evidence*, 13(4) *Applied Measurement in Education* 343, 351 (2000).) What matters in this context is less about labels, and more about conceptual clarity with respect to what criteria mean, how they are used, and the analysis that must accompany any validation determination.

<sup>17</sup> *Association of Mexican American Educators v. State of California*, 231 F.3d 572 (9<sup>th</sup> Cir. 2000) (en banc).

grade bilingual education teachers to ...librarians, and administrators” – the test was found to be valid in part because the “basic skills in reading, writing, and mathematics” were important elements of the jobs of public school teachers, across the board.<sup>18</sup> Thus, part of the apparent strength of California’s case involved the fact that the skills tested were considered so basic and fundamental to teaching that, despite technical problems with the validation process, it strained credulity not to require the skills measured by the test for any teacher in the California system.

In cases where the use of a high school diploma as a condition for employment has been challenged, courts have tended to accept the validity of the diploma in cases where the job in question was one involving public safety, high risk, or fairly indisputable complex skills. Notably, one federal circuit court has concluded that diploma requirements are less likely to be misused than employer-generated tests, and that a diploma requirement for a particular position might be sufficiently obvious and appropriate so as to permit “dispensing with empirical validation.”<sup>19</sup> By the same token, courts have tended to strike such requirements when they have been imposed as conditions related to low-skilled jobs and where evidence indicated that non-high school graduates performed well in the job in question.<sup>20</sup> Notably, some courts rejecting diploma requirements have indicated a willingness to consider course requirements that are more particularly linked to the job in question (e.g., a requirement of a year of algebra or geometry for positions in the sheet metal trade).<sup>21</sup>

---

<sup>18</sup> The record in the case reflects numerous technical problems with the validation process (including, as noted by the lower court, reliance on an “unscientific” and “not particularly helpful” study).

<sup>19</sup> *Aguilera v. Cook County Police and Corrections Merit Board*, 760 F.2d 844 (7<sup>th</sup> Cir. 1985) (ruling that high school education was an appropriate credential to require of corrections officers).

<sup>20</sup> See, e.g., *Griggs v. Duke Power*, 401 U.S. 424, 431 (1971) (ruling among other things that the high school completion requirement did not bear a “demonstrable relationship” to the successful performance of jobs for which it was a condition, and that no “meaningful study” of the relationship between the diploma requirement and job-performance had been conducted); see also *Watkins v. Scott Paper Co.*, 530 F. 2d 1159, 1182 (5<sup>th</sup> Cir.), *cert. denied*, 429 U.S. 861 (1976) (desire to upgrade the workforce is, standing alone, an insufficient basis for imposing high school diploma requirement).

<sup>21</sup> See, e.g., *Reynolds v. Sheet Metal Workers Local 102*, 498 F. Supp. 952 (D.D.C. 1980), *aff’d*, 702 F. 2d 221 (D.C. Cir. 1981) (rejecting high school diploma requirement but suggesting that specific course requirements might suffice for apprentice program); see also *United States v. Georgia Power Co.*, 474 F. 2d 906, 918 (5<sup>th</sup> Cir. 1973) (observing that high school diploma requirements included numerous courses that were not necessary for job-specific abilities, and rejecting the diploma requirement). Compare *Lombard v. School District*, 463 F. Supp. 566 (W.D. Pa. 1978), a case in which the court refused to compare the grades of candidates competing for a position of employment, concluding that the inherent *absence* of standards relating to grading practices—particularly at different schools at different times among different generations—would preclude meaningful reliance on those grades. That court also refused to accept the notion that

## Comparing K-12 Education, Higher Education, and Employment

In many ways, the due process and nondiscrimination analyses described above are similar in the K-12, higher education, and employment contexts, but some potential distinctions among these contexts (frequently observed by courts) are important to keep in mind:

- *The use of tests often serves different purposes in different contexts.* Tests of student achievement administered in the K-12 setting (such as most high school exit exams) are often backward-looking and designed to certify mastery of core content. Comparatively, tests used in the higher education and employment contexts (such as tests used in admissions, placement, or hiring decisions) are often forward-looking and designed to predict performance. These uses are different, and to some extent create different obligations regarding core issues such as the nature of validity evidence required and the extent to which “opportunity to learn” is essential to test validity.
- *Deference is often given to education judgments.* Based largely on principles of separation of powers and federalism, courts have repeatedly affirmed the importance of local control in education and that states have substantial authority to make educational decisions. Accordingly, courts tend to defer to state and local education judgments, recognizing that educational choices (especially with regard to goals and objectives) are peculiarly matters of state legislative policy and involve judgments about which courts lack expertise.<sup>22</sup> This is especially true in the context of K-12 education, but applies to some extent in the higher education context as well (particularly where issues of academic freedom may surface). Comparatively, decisions affecting employment opportunities may not merit such deference.
- *K-12 obligations are expansive.* At the same time, education, specifically at the K-12 level, implicates obligations broader than that of simply ensuring accurate measurements on tests or other criteria. In the words of one court:

---

important qualities such as leadership and maturity could be sufficiently measured by academic grades.

<sup>22</sup> See, e.g., *San Antonio Indep. Sch. Dist. v. Rodriguez*, 411 U.S. 1, 42 (1973); *GI Forum*, 87 F. Supp. 2d at 667-68, 670.

[I]f tests can predict that a person is going to be a poor employee, the employer can legitimately deny that person a job, but if tests suggest that a young child is probably going to be a poor student, the school cannot on that basis alone deny that child the opportunity to improve and develop the academic skills necessary to success in our society.<sup>23</sup>

Thus, the focus on opportunity to learn – including curricular and instructional alignment with assessments as well as targeted educational interventions to help students achieve to learning standards – is central in the K-12 setting, with the driving question being whether the educational practice at issue constitutes an educational benefit for students.<sup>24</sup>

Whether in the education or employment context, one important lesson that can be derived from the federal opinions is that a prudent, probing examination of the kinds of evidentiary questions courts will pose – as part of a sustained and systemic process in which stakeholders representing multiple disciplines are involved – can establish the kind of foundations that can minimize legal risk and exposure. The challenge is to ensure that all of the right (and frequently hard) questions are asked on the front end, rather than (for the first time) in a deposition or courtroom.

---

<sup>23</sup> *Larry P. v. Riles*, 793 F.2d 969 (9<sup>th</sup> Cir. 1984).

<sup>24</sup> See, e.g., *GI Forum*, 87 F. Supp. 2d at 681.

## **II. THE AMERICAN DIPLOMA PROJECT AND FEDERAL LAW**

The American Diploma Project (“the ADP”) seeks to strengthen on-going standards reform efforts in five states, and ultimately the nation, by bringing together each state’s K-12, higher education, and business communities to work more closely in support of high standards. The goals of the ADP are:

- To assist states in strengthening and/or revising their current standards in English language arts and mathematics to fully align high school graduation standards with college admissions and employment standards;
- To develop demand among colleges and employers for standards-based high school assessment data to be used in admissions and hiring decisions; and
- To develop new high school graduation benchmarks in English language arts and mathematics that can support future ADP-like efforts in other states.

The ADP provides prospects for advancing the standards reform movement in significant ways – helping ensure that all students receive access to the kind of education that can prepare them for success in college and/or the work force, as well as reducing the need for remedial education and training. As with the standards reform movement more generally, however, several legal issues are implicated by the ADP, specifically with regard to due process and nondiscrimination laws – highlighting the need for a careful and ongoing legal analysis that can help avoid unnecessary litigation and strengthen the ultimate defensibility of the project’s design and state-specific implementation – all while improving the educational outcomes. The bottom line is this: A delicate but discernible line exists between good and bad standards-reform policies and implementation strategies. The former hold great promise; the latter raise substantial risks and have the potential to subvert the very goals of the ADP.

### **Three Ways the ADP Implicates Federal Law**

The ADP implicates federal law in three primary ways, described below. In some cases, the ADP may raise new, novel legal issues. In other cases, the legal issues associated with the ADP may already exist within standards reform, below the litigation radar, but the ADP may “shine a light” on those issues in a way that raises their visibility. In either case, the legal issues need to be fully understood



and evaluated at the front end to ensure the most appropriate design and implementation of the ADP.

***Raising the Bar.*** First, the ADP's focus on aligning state high school graduation standards with college and employment standards suggests, in some cases, a movement toward higher standards in the K-12 context. In states that currently have high school exit exam requirements (or that attach other individual, high-stakes consequences to high school assessment data), this movement could *raise the bar* for a high school diploma (or other benefits). This would potentially implicate federal legal issues in two ways: Most notably, the establishment of new, higher academic requirements would likely implicate anew federal laws related to the use of assessments for such high-stakes purposes, including due process and nondiscrimination laws (to the extent that raising or altering the bar for a high school diploma creates or exacerbates disparities by race, national origin, or other protected classifications). It may also expand states' obligations with respect to the quantity and quality of validity evidence necessary to support the imposition of individual, high-stakes consequences set at those higher standards, including with regard to opportunity to learn requirements.

***Raising the Stakes.*** Second, the ADP's focus on encouraging colleges and employers to use standards-based high school assessment data (or diplomas linked to that data) in their admissions and hiring decisions, respectively, would *raise the stakes* for high-school assessments by tying additional benefits, and the denial of such benefits, to high-school assessment data. Most notably, this would create new sets of legal obligations, as the given assessment data would likely have to be independently valid for each specific use. Furthermore, as above, this may expand the states' (or others') obligations with regard to the quantity and quality of validity evidence necessary to support the imposition of individual, high-stakes consequences, because the extent of validity required is likely directly related to the extent of the consequences that attach to the given assessment. (It is important to note, however, that the increased use of the assessment data may not be creating new high-stakes decisions but rather may provide a different or additional mechanism for informing decisions that are already being made, and the relative validity of the decisions with or without the additional assessment data is, therefore, likely relevant to some degree.)

***Breaking Down Barriers.*** Third, the ADP's focus on bringing together the K-12, higher education, and business communities to fully align and use state criteria such as high school assessment data for additional purposes would have the effect of *breaking down barriers* in ways that might create new legal obligations for certain actors. In particular, along with the creation of a more express relationship between high school assessment data and college admissions, the potential linkages established between public higher education and K-12 systems may create or shine a light on the responsibility of states, and potentially public

universities, to ensure that students have a meaningful opportunity to learn the material being tested. (At the same time, breaking down these barriers can also help develop assessment systems that are more completely aligned and more likely to produce data that are valid for multiple purposes.)

It is important that states fully comprehend the three ways that the ADP potentially implicates federal law, and that states – including their K-12, higher education, and business communities – move forward in a way that can most effectively minimize legal risk.<sup>25</sup> A framework to guide that effort is provided in the next section.

---

<sup>25</sup> In addition, ADP states will have to address a host of practical issues that are beyond the scope of this paper. For example, will SAT, ACT, or similar test scores count in lieu of state-specific tests used for college admissions purposes? If so, will in-state students have the option of gaining admission without regard to the state-specific test? By contrast, will out-of-state students be required to take the state-specific test as a condition of college admission to state public institutions?

### III. A FRAMEWORK FOR IDENTIFICATION AND ANALYSIS OF CENTRAL ISSUES

Given the federal legal standards regarding due process and nondiscrimination that apply to the use of assessment data (or other factors) by states, colleges, or employers in making individual, “high-stakes” decisions, and given the several ways in which the American Diploma Project (“the ADP”) specifically implicates those federal legal standards, this section provides a framework to help the states participating in the ADP – including their K-12, higher education, and business communities – understand the relevant legal issues that may arise in the ADP context and move forward in a way that is legally appropriate and consistent with ADP goals. Simply put, every actor that will potentially be using state assessment data to make decisions regarding the provision of individual educational or other benefits, (e.g., graduation, admissions, or hiring decisions) should address four central questions derived from the primary legal standards described above:

- *Purpose.* What is the purpose of the assessment (or other criteria with consequences) in terms of how the assessment data are being used by the given state, university, or employer?
- *Validity.* Is the assessment valid for its intended purpose(s)?
- *Administration.* Is the assessment being administered in ways that ensure the validity of the decisions being made?
- *Opportunity to Learn.* Have students had a meaningful opportunity to learn the material being tested (to the extent necessary for the given purpose)?<sup>26</sup>

The chart below summarizes the analysis that follows. Each cell contains information based on reasoned assumptions about the ADP and its likely implementation. This framework is designed, in large part, to facilitate conversations among educators, policy makers, business leaders, and attorneys in each state – all of whom have important roles to play in shaping the state’s ADP efforts.

---

<sup>26</sup> The distinctions among issues explained in this framework should not mask the fact that, at important analytical junctures, the issues converge. So, in some contexts, for instance, it becomes impossible to disassociate accommodations issues from core validity questions, or timing questions from the issue of how many test administrations are allowed. Moreover, it should be noted that these categories have been established for the purposes of clarity and ease of reference, based generally on the way that courts tend to consider these issues. This organization is not intended to represent a framework that would fully track a psychometric discussion of the issues.

## A Framework for Analysis of ADP Federal Legal Issues

Key Issues	State/School District	College/University	Employer
<b>Purpose</b>	To certify mastery; to provide foundations for college admissions and employment decisions	To ensure admission and placement of students who will succeed in educational program	To ensure hiring of employees who will function effectively in their jobs
<b>Validity</b> <ul style="list-style-type: none"> <li>▪ Test development</li> <li>▪ Alignment</li> <li>▪ Cut scores</li> </ul>	Requires evidence that assessment is aligned with state standards	Requires evidence that assessment is aligned with factors related to (i.e., is predictive of) success in college; educational necessity	Requires evidence that assessment is aligned with factors related to (i.e., is predictive of) job success; business necessity
<b>Administration</b> <ul style="list-style-type: none"> <li>▪ Multiple opportunities</li> <li>▪ Multiple measures</li> <li>▪ Accommodations</li> </ul>	Requires use of assessments in a manner designed to avoid false negatives in measuring knowledge of state standards	Requires use of assessments in a manner designed to avoid false negatives in measuring ability to succeed in college; broader constructs implicated	Requires use of assessments in a manner designed to avoid false negatives in measuring ability to perform in job; broader constructs implicated
<b>Opportunity to Learn</b> <ul style="list-style-type: none"> <li>▪ Curriculum, instruction</li> <li>▪ Interventions</li> <li>▪ Data</li> <li>▪ Resources</li> <li>▪ Timing</li> </ul>	Requires that students have meaningful opportunity to learn the material being tested	May require that students have meaningful opportunity to learn material tested or that extent of opportunity is otherwise considered, at least with public institutions	Opportunity to learn not clearly implicated

## 1. What Is the Purpose of the Test?

This question seems simple and obvious, but it is all-too-often anything but. It implicates a series of complex and multi-faceted inquiries, ultimately centering on the following: Why is the state, college, or employer using the assessment data? What is the actor trying to measure? And what inferences are being made regarding the data?

Under federal law, courts tend to require only that the given purpose for using the assessment data be legitimate in the sense that it is reasonably related to the given actor's educational, business, or other interests. This is a relatively low threshold, especially in the education context where substantial deference is often given to states or colleges to define their educational interests (including such interests as improving the quality of education, ensuring that graduates are prepared to compete nationally, or ensuring that a high school diploma represents a particular level of achievement).<sup>27</sup>

However, the importance of this inquiry derives from the fact that this issue establishes the foundation and direction for the remainder of the legal analysis, regardless of the context. Only by understanding the purpose(s) of the given assessment in terms of how the data are being used by the particular actor – state, college, or employer – can it be determined if the assessment is valid for that purpose, and is, therefore, on solid legal ground. Moreover (as described in detail below), it is likely in the ADP context that the purposes for using state assessment data tied to new standards developed through the ADP process will differ among, and within, each state's K-12, higher education, and business communities. For example, while states with high school exit exam requirements may use the data to certify mastery of core content, colleges and employers may use the data to predict future performance – two very different purposes likely requiring different validity evidence. Thus, the legal analysis will necessarily differ to some degree as well.

Importantly, the establishment of an institution's purpose(s) involves more than mere statements. Courts will certainly consider an actor's articulated purpose for using assessment data, but courts will likely also seek to determine the actual purpose. And it is vital that there be alignment between the two, as well as evidentiary foundations supporting that alignment, for inconsistency between purported and actual goals may signal potential problems in validity.

Finally, this need for underlying consistency of purpose suggests that ADP participants' understanding and statements of test objectives must be defined with sufficient clarity to support meaningful analysis. For instance, a test may be

---

<sup>27</sup> See Arthur L. Coleman, *Excellence and Equity in Education: High Standards for High Stakes Tests*, 6 Va. J. Soc. Pol'y & L. 81, 95 (1998).

established under state law to “attract the best and the brightest” students within the state to its higher education institutions; another test may also be established to “help predict success in college” as measured by first-year grades, graduation rates, and the like. The former objective, which is inherently ambiguous and devoid of operational meaning, may without more fail to provide a baseline by which to evaluate validity on the front end, and may be subject to substantial legal debate regarding its meaning if challenged on the back end. The latter statement (or ones like it), which provides a baseline for substantive evaluation, can provide an important foundation for conducting such analysis.

## **2. Is the Assessment Valid for that Purpose?**

Ultimately, this inquiry centers upon the question of whether the test in fact measure what it says it measures: What inferences are being made about the meaning of the given test scores, and are those inferences likely accurate?

As a starting point, to reach legally sustainable conclusions to these questions in the ADP context, it is important to recognize that because different actors in each state – K-12, higher education, and business – will likely have different purposes for using the assessment data, each actor must have the evidentiary foundations necessary to show that the data are valid for its purpose.

In the K-12 context, this means, to the extent that there is a high school graduation exam requirement that seeks in part to certify mastery of core content, that the primary questions are likely (1) whether the assessment is aligned with state standards such that performance on the exam accurately measures student knowledge of the standards, and (2) whether the assessment is aligned with curriculum and instruction such that higher performance on the exam accurately measures student achievement and not opportunity to learn, or lack thereof. (To the extent that states move toward more forward-looking assessment goals, such as providing foundations for college admissions and employment decisions, additional validity issues may arise.) In the higher education context, where the assessment data will most likely be used to inform admissions or placement decisions, the primary question is likely whether higher performance on the assessment is correlated with greater success in the given college program, which could be defined in several ways based on college grades, retention, graduation, etc. In the employment context, the primary question is likely whether higher performance on the assessment is correlated with higher performance on the job.<sup>28</sup>

---

<sup>28</sup> To the extent that use of assessment data means reliance on specific cutscores, those cutscores must be validated in the sense that performance above the cutscore accurately reflects a qualitatively different level of knowledge or skills than performance below that cutscore. As a general proposition, federal law requires that cutscores be established based upon reasonable

In the ADP situation more broadly, however, several additional points (discussed in detail below) regarding the validation of test uses consistent with federal legal standards bear emphasis:

- *The same test may be validated for multiple purposes, with the right foundations.* Federal laws do not preclude the validation of one instrument for more than one purpose as long as it is independently valid for each purpose and the purposes do not conflict. However, one generally may not impute a conclusion about validity for one purpose when addressing a test use for a different purpose.
- *In the right circumstances, validation efforts may occur during the implementation of the testing program.* The perfect does not have to be the enemy of the good. With appropriate foundations and the right balance, additional validation efforts may likely be pursued even as the tests in question are being used – but potentially with greater legal risk.
- *Process matters.* In examining test validity, courts frequently examine the actual process of validation, with a focus on factors such as consistency with psychometric standards and attention to inclusiveness among appropriate actors.

**Validation for Multiple Purposes.** Whether in the educational or employment context, federal law generally requires that a test that has a discriminatory impact or that implicates due process be valid for its particular purpose. As a broad proposition, nothing in federal law prohibits multiple uses of a single test or instrument for related (or for that matter, differing) purposes – as long as the test is independently valid for each purpose. The question is at its core one of psychometrics: If there is a sufficient scientific basis upon which to conclude that one instrument may be used for different purposes, federal courts are unlikely to

---

conclusions, reflecting the exercise of good faith professional judgments informed by facts and data generated through a systemic process. See, e.g., *GI Forum*, 87 F. Supp. 2d at 680; *Association of Mexican-American Educators v. California*, 937 F. Supp. 1397 (N.D. Cal. 1996), *aff'd*, 231 F.3d 572 (9<sup>th</sup> Cir. 2000) (en banc); *Cureton v. NCAA*, 37 F.Supp. 2d 687 (E.D. Pa. 1999), *rev'd on other grounds*, 198 F.2d 107 (3<sup>rd</sup> Cir. 1999). In essence, although the basic rules related to cutscores are sometimes more relaxed than those with respect to validity issues in general, the answers to the questions with respect to general validity provide a useful roadmap for making decisions about cutscores. For instance, multiple ADP actors may rely on the same assessment but may possibly judge “merit” in different ways (such as where a score of 70 on a high school exit exam indicates sufficient mastery of material to constitute a pass but where that same score, considered in the context of other admissions criteria, would not qualify a student for admission to a university). Such judgments may be legally sustainable so long as there are the necessary supporting evidentiary (e.g., psychometric) foundations justifying the distinctions and there is a fundamental harmony among the actors’ decisions regarding cutscores.

second-guess that result (absent, of course, competing and stronger psychometric testimony to the contrary).<sup>29</sup>

Potential conflicts arise, however, when the chain of inferences necessary for a test to be valid for two purposes appear to contradict each other. The case that has most often been raised, though it is very much an open issue, is where a test that is used for purposes of *school* accountability – in which low scores are used to indicate that a school is “failing” – is also used for individual *student* accountability – in which a low score is used to indicate that a student in a low-performing school has not learned material that he/she potentially was never taught.

This circumstance does not necessarily seem to be implicated by the ADP. Broadly speaking, a test can potentially be valid both as a measure of what students have learned and how they are likely to do in college or the work force. In fact, some empirical basis already exists in support of that proposition.<sup>30</sup> At the same time, given the potential tension, perceived or real, that may exist with respect to using the same test for multiple purposes, caution is appropriate – particularly in the event that there are circumstances in which the tension or inconsistency among multiple purposes would make such an effort improbable or impossible.<sup>31</sup>

In the ADP context in particular, the goal of having multiple actors rely on a single assessment for multiple purposes makes it advisable for all relevant actors to arrive at an agreed set of testing (and related) objectives and constructs – aligned with their distinct purposes. Notably, the challenge of establishing with clarity this baseline common ground is not without benefits. If colleges and employers participating in the ADP properly identify what standards are important to success in college or on the job, and if those standards are accurately reflected in state high school education standards, then assessments that accurately reflect student achievement relative to those standards may have a strong foundation of validity for multiple purposes.

---

<sup>29</sup> See generally OCR *Resource Guide* at 55-56; Coleman, *Excellence and Equity*, *supra* note 27, at 101, 103-04; *Joint Standards*, Standard 13.2 (“In educational settings, when a test is designed or used to serve multiple purposes, evidence of that test’s technical quality should be provided for each purpose.”).

<sup>30</sup> E.g., ACT, Inc., *Content Validity Evidence in Support of ACT’s Educational Achievement Tests*. (2000).

<sup>31</sup> See generally *Joint Standards*, Standard 13.2 (“No test will serve all purposes equally well. Choices in test development and evaluation that enhance validity for one purpose may diminish validity for other purposes. Different purposes require somewhat different kinds of technical evidence....”).



Finally, the potential for multiple parties in ADP states to establish standards in coordination and use the same assessment data for related purposes sets the stage for possible implementation strategies that could merit significant court deference regarding the nature and extent of validity evidence that is required (even in the context of higher stakes). This premise is suggested by an analysis of the *GI Forum*, *AMAE* and *Aguilera* cases, which, although addressing factually distinguishable test use challenges, together hold out the prospect that a test associated with basic (albeit high-standards) knowledge or skills in the ADP context could more likely satisfy legal validation standards where some combination of the following factors is present:<sup>32</sup>

- A history or tradition of reliance on a (similar) set of common-sense indicators as a foundation for relevant decision-making exists;<sup>33</sup>
- The test does not purport to define the universe of skills or abilities at issue in the decision-making process, but merely captures some subset of the relevant knowledge or skills;<sup>34</sup>
- What is being measured is deemed fundamental to a common core of basic skills (such as reading, writing, and math);<sup>35</sup>
- The indicator(s) relied upon are degrees awarded by schools or systems, rather than tests designed by employers;<sup>36</sup> and

---

<sup>32</sup> Although this list begs certain contextual questions that are central to any federal court decision, it nonetheless presents a possible outline that can help inform the ADP-related analysis in certain jurisdictions.

<sup>33</sup> See *Aguilera*, 760 F. 2d at 847 (ruling that validation was not required for the use of a high school diploma requirement for policeman or corrections officer, and observing that “[n]o one would insist that a law school validate statistically the ‘business need’ behind requiring that its faculty members have law degrees...or that a hospital validate a requirement that its doctors have medical degrees”) (internal citations omitted).

<sup>34</sup> See *AMAE*, 231 F.3d at 587 (ruling that job relatedness was sufficiently established on a test designed to measure “a minimum level of competence” where the test in question is “not intended to measure all the skills that are relevant to all of the jobs for which it is required [, or, for that matter] *all* of the skills of *any* of the jobs for which it is required”) (emphasis in original).

<sup>35</sup> See *GI Forum*, 87 F. Supp. 2d at 669, 681 (upholding test that measures “essential skills and knowledge,” an area of educational judgment in which the court lacks authority to interfere). See also *AMAE*, 231 F.3d at 585-86 (upholding test that measures “basic skills in reading, writing, and mathematics” as “important elements of work behavior” for public school officials) (conclusion by inference).

<sup>36</sup> See *Aguilera*, 760 F.2d at 847 (stating that “[t]ests ...made and scored by the employer...are easily misused; degrees ...awarded by schools that are independent of the employers who use the degrees as job qualifications [are not]”) (internal citation omitted).

- The use of the test is inextricably tied to projected educational benefits associated with the very system of which the test or instrument is a part.<sup>37</sup>

In *GI Forum*, for instance, the court, which ultimately rejected due process and discrimination challenges against the State of Texas's high school exit exam, reasoned (among other things) that the test in question was a measure of standards that the State "has determined are essential skills and knowledge," and that resulted in a "positive" impact on students statewide. Although not perfect, the court recognized that the State's system of accountability (of which the statewide test was central) reflected decisions that were within the State's authority: Texas—not the court—had the "right to pursue educational policies that it legitimately believe[d were] in the best interests of Texas students." Similarly, in the employment context, the Seventh Circuit in *Aguilera* observed that "sometimes the appropriateness of an educational requirement is sufficiently obvious to allow dispensing with empirical validation," citing to numerous opinions supporting the proposition that such validation was not necessary in the context of law schools requiring law degrees of their professors; hospitals requiring medical degrees of their doctors; of public health workers having college degrees; of professors having doctoral degrees; of police officers having a high school degree; or of certain employees knowing English.<sup>38</sup> In a related vein, the Ninth Circuit in *AMAE* upheld a teacher certification test that measured "basic skills in reading, writing and mathematics," finding that even though the validity foundations were less than "overwhelming,"<sup>39</sup> the test required a basic level of mastery that could be required of all teachers and did not purport to measure all job-related skills.

***Timing of Validation and Implementation.*** As indicated above, federal law requires evidence of validity whenever assessment data are used to make individual student, high-stakes decisions. Thus, it is important that the implementation and use of assessment data not get inappropriately ahead of efforts to ensure validity, including opportunity to learn (discussed below). However, recognizing that standards-based assessment systems can and should be educationally beneficial, federal courts have indicated that the perfect need not be the enemy of the good. Recalling that the question of whether a test is valid for a particular purpose is less a question that involves definitive lines than one

---

<sup>37</sup> See *GI Forum*, 87 F. Supp. 2d at 674, 683 (observing the remediation benefits that tied to the testing program and recognizing the deference afforded to states that "design an educational system that [they believe] best meets the needs" of their citizens).

<sup>38</sup> *Aguilera*, 760 F. 2d at 847-48.

<sup>39</sup> *AMAE*, 231 F.3d at 589.

implicating movement along a continuum, and that no test is in legal or psychometric terms ever held to the standard of perfection, educationally-grounded judgments along the way can be made about the level of evidence that (at any particular time) supports prospective use, or multiple uses, of a given instrument.

In the ADP context, the process by which each state's K-12, higher education, and business communities come together to set common standards may, in itself, produce some foundations for the validity of assessments developed through that process for multiple purposes. Moreover, in some states, initial validity analysis can likely be done based on existing data from high school exit exams, for example. Thus, while state-specific analysis is likely important here, ADP participants may be justified in moving toward the gradual, appropriate use of state assessment data (e.g., as one of multiple measures, as described below) even as efforts to enhance validity and opportunity continue. It is likely a question of balance – tying the extent of validity to the nature of the use. In this context, it may be worth reiterating that (at least in many states) the ADP is not creating new high-stakes decisions, but is trying to enhance the meaning and validity of existing high-stakes decisions.

***The Validation Process.*** The validation process, in addition to the validity evidence itself, may be of importance in building the foundation to legally support a given testing program and use of assessment data. Recognizing the importance of adherence to generally accepted psychometric principles, courts have looked to the validation process related to an assessment system in determining validity, rebutting claims of discrimination, and supporting core educational objectives. For example, cases in the educational and employment settings have emphasized the importance of ensuring that the process of test use validation involve diverse groups of experts providing input about the validity of the instrument. Local educator input has been judged to be critical in the high school exit exam setting, and the existence of a cultural and racial bias review can be important as part of that exercise. Moreover, as part of the field testing of any instrument, the practice of reviewing questions that have a disproportionate impact on students, based on race, national origin, sex, or disability – and making conscious decisions about whether to retain or discard the question as potentially discriminatory and unjustified – can be an important foundation for helping ensure a court's receptivity to the instrument.<sup>40</sup>

---

<sup>40</sup> See generally *GI Forum*, 87 F. Supp. 2d at 672.

### 3. Is the Test Being Administered and Used Appropriately?

This question is inextricably connected with broader issues regarding the validity of an assessment for a particular use. An assessment that is generally valid for a given purpose may not be valid, or as valid as it could be, for a particular individual in a particular case, depending on how the data are used. The legal requirements here, in essence, seeks to avoid false negatives with regard to decisions with high-stakes consequences, to the extent possible. Key issues (discussed in detail below) include:

- *Whether students have multiple opportunities to take the assessment,*
- *Whether multiple measures are used in making high-stakes decisions; and*
- *Whether appropriate accommodations have been afforded students with disabilities and English language learners.*

**Multiple Opportunities.** Most states in the K-12 setting permit several administrations of high school exit exams, providing instructional interventions for students who have not passed the test – a critical element in any opportunity to learn analysis (discussed below). These multiple opportunities are a core requirement for validity in the K-12 context as they help avoid false negatives, such as where a student simply had a “bad day,” and help ensure that the testing program is educationally beneficial.<sup>41</sup> The historical practice with respect to university or business decisions may be different in some cases, such as where test scores from more than one administration are averaged or considered in combination as part of admissions or hiring decisions. Given the ways that the ADP breaks down barriers among the K-12, higher education, and business communities, and the fact that colleges and employers would most likely be using state assessment data to predict student performance beginning *after* graduation, caution should be exercised if any effort is made to depart from the traditional state multiple opportunities model in using state assessment data.

**Multiple Measures.**<sup>42</sup> The relative weight placed on a single criterion such as a test score when making high-stakes decisions also is central to the ADP

---

<sup>41</sup> See *id.* at 675 (stating that the number of testing opportunities “limits the possibility of false, negative results”).

<sup>42</sup> There are few terms in testing that are subject to more debate and confusion than “multiple measures.” The confusion stems from the fact that the term, standing alone, fails to describe with particularity the practice in question. Therefore, for the purposes of this paper, the term “multiple measures” refers to the compensatory model in which a “student’s strong performance on one indicator, such as course work, [may] offset or compensate for low performance on another, such as the graduation exam.” NRC *High Stakes* at 180. This practice should be distinguished from the

discussion. An axiomatic principle of psychometrics is that no single test score can be considered a definitive measure of an individual's knowledge, skill, or ability.<sup>43</sup> Thus, to the extent that multiple, valid measures are considered when making a decision with consequences affecting individuals, there is an increased likelihood that the given decision is valid, with the legal exposure or risk correspondingly reduced.<sup>44</sup>

Multiple measures may not be legally required in all circumstances – most notably in cases in which tests are designed to measure student achievement or mastery on a narrow set of constructs. For example, consistent with *GI Forum*, though representing only one case on point, a state may decide that, no matter what other material is taught in school, a high school diploma is going to embody demonstration of mastery on some limited number of constructs covered on a given assessment, and that, irrespective of other factors, students must pass that test in order to graduate from high school – though such use of tests likely heightens legal risk.<sup>45</sup>

However, the federal court's approval of the use of a test as a "gatekeeper" in *GI Forum* does not easily lead to the conclusion that higher education officials or employers who use such instruments to gauge future success may do the same under federal law. In fact, given that no state assessment of English language arts or mathematics, as contemplated in the ADP, is likely to measure anything close to the full breath of information relevant to predicting future student performance in college or on the job, the use of multiple measures (e.g., grades, test scores, essays, recommendations, professional experience, etc.) when making, admissions, hiring, or other decisions in the higher education and business contexts is likely critical (and is likely consistent with current university and business practices).

---

conjunctive model, pursuant to which students are required "to complete all of their coursework satisfactorily *and* to pass the... test(s)." *Id.* This was the model used by the State of Texas in the administration of its high school exit exam. See *GI Forum*, 87 F. Supp. 2d at 675.

<sup>43</sup> See, e.g., NRC *High Stakes* at 180.

<sup>44</sup> Given the "raising the stakes" implications of the ADP, consideration should be given to the prospect that the use of multiple measures (where not previously used, for instance) might mitigate some of the legal risk otherwise implicated in the ADP implementation, particularly in the early stages.

<sup>45</sup> Similarly, OCR case resolutions regarding standards reform issues and high school exit exams have not imposed a multiple measures requirement on states. E.g., *State of Ohio*, OCR Case No. 15-94-5003; *Texas Education Agency*, OCR Case No. 06-96-1021; *State of North Carolina*, OCR Case No. 11-98-1070.

**Accommodations.** States, school districts, universities, and employers are required under federal law to offer appropriate accommodations in assessments for individuals with disabilities and, specifically with regard to states and school districts, for English language learners as well. In either case, the purpose of the given accommodations is to ensure that assessments used to measure the knowledge and skills of students with disabilities or limited English proficiency accurately measure their knowledge and skills on the desired constructs, rather than measuring irrelevant factors related to their disabilities or limited English skills (unless those are the skills that are purposely being tested). Given the federal legal requirements regarding meaningful inclusion of students with disabilities and English language learners in state testing programs, and the benefits that may flow from such inclusion, it is likely important that issues of accommodations be addressed at the early stages of any assessment program.

#### **4. Have Students Been Given a Meaningful Opportunity to Learn?**

The opportunity to learn issue, which is closely associated with validity determinations, is not traditionally implicated in all high-stakes contexts. It has generally surfaced in the K-12 setting where, in systems of compulsory education, states and school districts have undertaken to establish criteria of student accountability associated with their classroom learning. Most notably, opportunity to learn is implicated under federal due process and nondiscrimination laws in cases involving high school exit exams that seek to certify student achievement on state standards, as opposed to predicting future performance. Where low scores reflect only a lack of opportunity to learn, it is likely not valid, and “fundamentally unfair,” to hold students accountable in that regard.<sup>46</sup>

Given the potential of the ADP to “raise the bar,” “raise the stakes,” and “break down barriers” among each state’s K-12, higher education, and business communities, however, additional opportunity to learn issues may arise. In particular, states may incur increased responsibilities with respect to the establishment of opportunity to learn foundations, and additional actors may be implicated with regard to opportunity to learn issues, though in potentially limited ways. Some examples of analogous cases already exist. For example, a recent case was filed against the State of California, *Daniel v. California*, challenging the State’s failure to equally and adequately provide advanced placement courses to black and Hispanic students.<sup>47</sup> At the same time, a second case was recently filed against the University of California, now titled *Castaneda v. Regents of the University of California*, challenging under the Title VI disparate impact standard

---

<sup>46</sup> See, e.g., *Debra P.*, 644 F.2d at 397.

<sup>47</sup> See *Daniel v. Cal.*, No. \_\_\_\_ (Super. Ct. filed July 27, 1999).

the consideration of advanced placement courses as part of university admissions, in part because black and Hispanic students do not have equal access to such courses at the high school level.<sup>48</sup>

While there is a dearth of law that would provide clear guidance with regard to the unique opportunity to learn issues implicated by the ADP, existing opportunity to learn principles suggest the need to consider the following points: First, to the extent that any implementation of the ADP results in a state “raising the bar” for a high school diploma or “raising the stakes” by directly tying additional high-stakes consequences to state assessment data, the state’s obligation to establish the necessary opportunity to learn foundations may be expanded – as to both the quality and amount of effort and evidence necessary to support its student accountability practices.<sup>49</sup> Second, the ADP’s potential use of state assessment data for college admissions and/or employment purposes may “shine a light” on opportunity to learn issues to the extent that a student’s opportunity to learn, or lack thereof, may affect the validity of individual decisions, such as those regarding admissions or hiring.<sup>50</sup> This may raise federal legal issues in some circumstances, such as where the underlying lack of opportunity is correlated with race or ethnicity. Third, to the extent that the ADP “breaks down barriers” and brings together multiple actors to establish standards and assessments that measure student achievement in K-12 education, additional actors, particularly public colleges or universities, may bear some opportunity to learn responsibilities that have historically belonged exclusively to states or districts. At a minimum, the higher education and business communities should be committed to supporting states in meeting their K-12 opportunity to learn obligations because, in the ADP context, any legal or educational attacks on a state’s K-12 assessment system likely has substantial implications for the broader use of state assessment data by colleges or businesses. There is likely good news here as well: To the extent that a college or employer is using state assessment data as part of a broad,

---

<sup>48</sup> See *Castaneda v. Regents of the Univ. of Cal.*, No. 99-CV-525 (N.D. Cal. filed Feb. 2, 1999).

<sup>49</sup> The opportunity to learn standard emanates in part from due process principles, which as a general proposition affirm that the greater the potential harm or deprivation to an individual, the greater the due process responsibilities.

<sup>50</sup> Imagine, for instance, a case where two students receive the same score (75 out of 100) on a state assessment that is being used by a college in admissions to help predict future performance. One student comes from a high-income, high-quality school where no other student scored below 75; another student comes from a low-income, low-quality school where no other student scored above 75. Same scores, but the differences in opportunity to learn may result in different conclusions regarding the likelihood that each student will excel in a rich college environment. In other words, opportunity to learn may affect the validity of these individual decisions.

educationally beneficial system of standards reform, that is potentially important to sustaining the use of the data under federal law.<sup>51</sup>

Given these possible developments, it is important to keep in mind some central elements of the opportunity to learn requirement. The focus of the few (but consistent) court opinions on the subject (which span more than two decades) has been on the following factors:

- The alignment among the curriculum, instruction, and the high-stakes assessment;
- The timing of the attachment of high-stakes consequences to the assessment, both with regard to fair notice and the ability to ensure some level of alignment before high-stakes attach to a test; and
- The use of data and the establishment of educational interventions based on the data that take place over time to ensure all students are helped to meet their potential.<sup>52</sup>

## Lessons Learned

The framework and legal analysis presented above suggest several lessons for ADP state participants. Because of the nature of prior cases, most of these lessons apply most directly to states at this stage of the ADP, but some are applicable in the higher education and business contexts as well, and will become more so as the ADP process moves forward. The following, in no particular order, are ten such lessons that apply in the ADP context:

- *Recognize that individual decisions regarding the delivery or denial of educational benefits can implicate a range of federal laws.* The attachment of “high-stakes” consequences has generally, though not always, been the foundation for litigation claims under the federal laws discussed above. Where an assessment is used for an individual, high-stakes purpose as contemplated in the ADP – whether for graduation, admission, placement, or hiring – and that use results in the denial of some educational or employment opportunity, then federal laws apply and require, in broad terms, that the given assessment be valid and appropriate for its purpose.

---

<sup>51</sup> See, e.g., *GI Forum*, 87 F. Supp. 2d at 681-84.

<sup>52</sup> In a panel presentation in 2000 regarding the defense of the TAAS by Texas, one attorney representing the State of Texas stated that the “secret” to the success of Texas in the lawsuit was “data, data, data.”



- *Be aware that federal civil rights and other laws reinforce sound educational practices.* The more educationally sound a given assessment system – and the more focused it is on improving rather than simply measuring student achievement – the more likely it is to be consistent with federal legal requirements. Thus, in the ADP context, significant steps toward legal compliance can be taken through careful strategic planning and analysis.
- *Know your purpose.* The purpose, sometimes multiple purposes, for which a given actor is or will potentially use assessment data drives all other inquiry, legal and educational, regarding the validity of the assessment for that purpose. In the ADP context, different actors will likely have different purposes for using state assessment data, resulting in somewhat different analyses.
- *Establish reasonable standards for accountability.* Any standards reform effort is a “work in progress,” and federal courts are likely to recognize this. Set the bar where appropriate for the given goals, but avoid setting the bar so high that it bears little relation to the kind and quality of teaching and learning taking place, and unduly invites backlash (and potential litigation).
- *Gather validity evidence.* Any actor using assessment data should have a process for gathering evidence regarding whether the given test is valid for its specific purpose(s) and for enhancing validity over time. A test can potentially be used for more than one purpose as long as it is valid for each purpose and those purposes do not conflict. Moreover, validity evidence likely can (and should) continue to be amassed during implementation, as long as a balance is maintained and test use, specifically the imposition of “high-stakes” consequences, does not get too far ahead of test validity.
- *Promote inclusion and appropriate accommodations.* Students with disabilities and English language learners generally must be included in assessment programs with appropriate accommodations, and these issues should likely be addressed at the front end of the testing program.
- *Ensure appropriate use of assessment data.* The appropriate use of assessment data (e.g., promoting multiple opportunities tied to educational interventions and, in many cases, using multiple measures to evaluate performance) in making individual decisions can help avoid false negatives and result in a stronger position educationally and legally.

- *Have good data and use it to improve educational outcomes.* The importance of having systems that accurately reflect student performance in a manner that is disaggregated according to race, national origin, sex, disability, English language learner status, free and reduced price lunch status, and the like cannot be overstated. A good data system can be an important foundation for demonstrating a system-wide commitment to accurate results and success for all students.<sup>53</sup> Ensure that systems are designed to identify and address gaps in learning. What a state does with its picture of student performance – the standards for intervention and the support for remediation, for instance – are critical in establishing that the right foundations are in place so that students can *fairly* be held “accountable.”<sup>54</sup>
- *Focus on inputs—early and often.* Ensure that there is a coherent statewide system in place (formal or not) for ensuring that common curricular and instructional foundations (i.e., opportunity to learn) exist for all students, both at the front end and through interventions and remediation at the back end.<sup>55</sup>
- *Focus on timing.* The establishment of validity, including curricular validity, and the attachment of high-stakes consequences should be balanced. The perfect need not be the enemy of the good, but neither should the stakes get too far ahead of the information and conditions that support sound judgments. It is a question of balance that suggests a gradual approach, with substantial room for variation.

---

<sup>53</sup> See generally No Child Left Behind Act of 2001, Pub. L. 107-110, 115 Stat. 1425 (2001) (imposing federal requirements of this sort for the purposes of state or district accountability).

<sup>54</sup> E.g., *GI Forum*, 87 F. Supp. 2d at 673-74.

<sup>55</sup> See *Debra P.*, 730 F. 2d at 1416-17 (upholding use of high school exit exam in state without state mandated curriculum but where significant state intervention had occurred to provide students with the opportunity to learn the material tested); *GI Forum*, 87 F. Supp. 2d at 682 (upholding use of high school exit exam in state where state mandated curriculum was complemented by significant state interventions designed to provide students with the opportunity to learn).

## IV. THE PROCESS FOR MOVING FORWARD

The framework presented above provides a list of questions and lessons learned that apply to each ADP state. How these questions and lessons apply depends to some degree on each state's legal and policy context. Broadly speaking, the next step is for each state to apply this guidance to its specific state context, and to develop a state-specific plan to ensure that the key legal issues discussed above are addressed throughout the ADP's development and implementation process. Said differently, the players in each ADP state who are using or may ultimately use state assessment data as part of individual, "high-stakes" decisions – including the K-12, higher education, and business communities – should come together and actively address the issues identified above.

Moreover, as the ADP process moves forward, there are several steps that each state can take to ensure that the issues identified above are addressed as part of the broader ADP effort. This is not intended to be an exhaustive list or a detailed explanation of each step. Rather, it is intended to provide a broad shell, in appropriate order, within which more specific steps and state-specific context can be added.

- ❑ **Identify key college and employment standards.** First, it is important for the higher education and business communities to each come together and identify the standards that are likely important to them for admissions and hiring purposes, respectively. Importantly, what standards are identified should be driven by each community's likely use of the data regarding students' knowledge of those standards. In other words, to the extent possible, the college and employment standards identified should likely be related to predicting success in college or on the job. In this way, efforts to align state standards will more likely lead to student assessment data that are valid for use in college admissions and hiring practices.
- ❑ **Conduct a "gap analysis" and align state standards to the extent appropriate.** Second, it is important to bring together the K-12, higher education, and business communities to review the state high school education standards and determine where there are gaps between those standards and the college and employment standards (identified above). The focus should then be on aligning K-12 standards with the higher education and business standards to the extent appropriate. (There should likely be room for discussion here regarding the lines that are being drawn in terms of student learning. In other words, where does high school learning properly end and college learning or job training properly begin?)

- ❑ **Develop a plan for ensuring that state curriculum and instruction, as well as educational interventions, are aligned with the new state standards.** Third, it is important to focus, early and often, on the high school education program and its alignment with any revised state standards. This is especially key for states with an existing high school exit exam requirement to the extent that the new standards are going to be phased in as part of that requirement. Efforts to align curriculum, instruction, and other resources to state standards can help ensure opportunity to learn on the front end. Moreover, if and when assessment systems are in place, the establishment of comprehensive data collection systems (disaggregated by race, ethnicity, etc.) can be crucial to identifying achievement gaps and promoting educational interventions on the back end to ensure that all students have a meaningful opportunity to learn the material being tested.
- ❑ **Develop a plan for assessing student knowledge of new state standards and ensuring the validity of assessments.** Fourth, each state will need to identify how it is going to measure student achievement on revised state standards and provide data that colleges and employers may potentially use in their admissions or hiring decisions. This will likely involve assessments of student achievement (though it could include other factors such as successful completion of certain courses of study in high school). For example, a state may decide to incorporate the new standards into high school exit exams. In this case, the exams need to be modified and validated over time. Comprehensive and inclusive validation processes should be established to gather and analyze information related to each potential use of the assessment data. States also need to focus on issues of test administration, such as multiple opportunities and accommodations for students with disabilities.
- ❑ **Develop a plan to phase in the appropriate use of high school assessment data in college admissions and hiring decisions.** Fifth, each state could develop a plan as part of the ADP with regard to the appropriate use of high school assessment data in admissions and hiring decisions. How quickly this move should be made likely depends on several factors, including findings from the gap analysis, the manner in which the data are used, the evidence that the data are valid for the given purpose(s), and more.